

Information Entropy Dynamics and Maximum Entropy Production Principle

Alexander L. Fradkov, *Fellow, IEEE*, Dmitry S. Shalymov

Abstract

The asymptotic convergence of probability density function (pdf) and convergence of differential entropy are examined for the non-stationary processes that follow the maximum entropy principle (MaxEnt) and maximum entropy production principle (MEPP). Asymptotic convergence of pdf provides new justification of MEPP while convergence of differential entropy is important in asymptotic analysis of communication systems. A set of equations describing the dynamics of pdf under mass conservation and energy conservation constraints is derived. It is shown that for pdfs with compact carrier the limit pdf is unique and can be obtained from Jaynes's MaxEnt principle.

I. INTRODUCTION

Matters of evolution in the nature and society raise a great number of various discussions. One of the efficient ways to describe and interpret the evolution is to use entropy.

Entropy is an information theory tool, designed to quantify the information content and possibly information loss for various systems. The term of information can be used as an inverse of uncertainty. In this case the growth of uncertainty is identified with an increase of the entropy which in turn is interpreted as the information loss.

Entropy and related information measures provide useful descriptions of the long term behavior of random processes. The notion of entropy often becomes the center of discussions both in statistical physics and in thermodynamics.

To describe the evolution of non-stationary processes that follow the principle of maximum entropy the maximum entropy production principle (MEPP) is often used [1]. The MEPP is that a system evolves in “natural” environment maximizing its entropy production when some restrictions are imposed. The “natural” environment means that there is no direct external action. Thus if the principle of maximum entropy (MaxEnt) says that a system tends to its state of maximum entropy then the MEPP says that the system tends to it at the highest possible rate. In some cases of system evolution analysis, these principles allow one to obtain missing information required in order to determine the direction of further evolution.

For now, attempts to justify MEPP widely use an approach based on the information entropy proposed by E.T. Jaynes (1957) [2]–[4]. This approach is a simple and convenient way to construct (both classical and quantum) statistical thermodynamics; it does not require some complexities such as the ergodic

hypothesis. An important advantage of Jaynes' formalism is its ability of being generalized to analyze non-equilibrium systems.

An approach to derive the MEPP using Jaynes' formalism is proposed by Dewar [5], [6]. This approach seeks the most probable path of evolution which describes non-equilibrium stationary systems. Dewar analyzes a non-equilibrium stationary open system (with volume V and boundary Ω) and uses a concept of "microscopic path" that implies microstate changes over time. However, according to the discussion in [1], it is still an open problem to trace connection between the maximum information entropy of trajectories and the MEPP, and therefore to determine the entropy evolution of the system. We propose a new justification of MEPP based on control theory.

It is known that methods of the optimal control (such as Bellman dynamic programming, Pontryagin's maximum principle, etc.) can be effectively used to develop models of mechanical [7], thermodynamic [8] and other complex systems. Using such methods also seems promising in the context of the MEPP. Equations of the dynamics of physical systems can be obtained from extremal principles if the goal is to reach the extremum of an objective functional. We can consider the entropy maximization of a system as such a goal and use the speed-gradient (SG) principle [9]–[13] originated in control theory. In this case, having determined the equations of the system dynamics that ensure the maximum entropy increment, the SG principle can justify the MEPP.

The SG principle is successfully applied in [9], [14] to obtain equations of the statistical dynamics of finite systems of particles that follow the principle of maximum entropy. Applicability of the SG principle is experimentally tested in [12]; it is done on the system of particles simulated with the molecular dynamics method on the basis of equations of classical mechanics. However, it is still an open question to apply these results to systems with continuous distributions.

An approach proposed in this paper provides the evolution law of the system in the following form:

$$\dot{p}(t, r) = -\Gamma(I - \Psi) \log p, \quad (1)$$

where I is identity operator, Ψ is a linear integral operator that is invariant for p , $\Gamma > 0$ is a constant gain.

By means of this approach the distribution corresponding to the maximum value of entropy can be found. A way how to find the distribution achieving the extreme value (maximum or minimum) of entropy within a given variational distance from any given distribution is proposed in [15]. Approximation of probability distribution is built there based on a new bounds on entropy distance. Also these bounds are applied to entropy estimation. The application of these bounds to typical sequences is discussed in [16].

Convergence of differential entropies which corresponds to pdfs from eq. (1) can be proven on the basis of sufficient conditions provided by Godavarti and Hero [17]. According to Godavarti and Hero

[17], the convergence of differential entropies can be traced to the problem of asymptotic analysis of various communication systems where the maximum asymptotic rate of communication is investigated [18]–[22]. Also the differential entropy convergence can be used in communication systems where the asymptotically optimal source compression [23], [24] or the asymptotic storage capacity [25] is analyzed. Entropy estimation and the convergence of differential entropies are also investigated in [26]–[28]

Jaynes’s MaxEnt principle was successfully used for inductive inference in [29]. It is shown that given information in the form of constraints on expected values, there is only one appropriate distribution which can be obtained by maximizing entropy. Appropriateness of this unique distribution is that it satisfies the constraints that can be chosen by a procedure that satisfies the consistency axioms. Continuous probability densities are also considered there. Inductive inference can be used in systems with non-stationary processes which satisfy the MaxEnt [31]. In [31] the notion of Entropy Dynamics (ED) is used. ED is a theoretical framework which combines inductive inference [29] and Information Geometry [30]. ED investigates the possibility of deriving dynamics from purely entropic arguments.

The main result of this paper is the extension of the [9], [14] results to systems with continuous probability distribution. The SG principle is used to derive equations of the dynamics of transient states of the systems that in a steady state follow the principle of maximum entropy.

The next section formulates the SG principle. Jaynes’ formalism is introduced in 3rd section. The 4th section gives an example of a dynamic system with continuous distribution of parameters that follows the principle of maximum entropy. Cases with one and two constraints are reviewed. Equations of the transient state are derived; and their properties are analyzed. The asymptotic stability of pdfs and differential entropy convergence is proved.

II. SPEED-GRADIENT PRINCIPLE

Let us consider a category of open physical systems which dynamics is described by the system of differential equations

$$\dot{x} = f(x, u, t), \quad (2)$$

where $x \in \mathbb{C}^n$ is the system state vector, u is the vector of input (free) variables $t \geq 0$. The task of system simulation (model development) can be defined as deriving the law of variation (of evolution) $u(t)$ that satisfies some criterion of “naturalness” of its performance to give the model features characterizing a real physical system.

Developing such a criterion based on extreme and variational principles usually implies specifying some integral functional (for example, an action functional of the principle of least action [32]) that characterizes the system performance. Functional minimization defines probable trajectories of the system $\{x(t), u(t)\}$

as points in the corresponding functional space. The advanced apparatus of the calculus of variations is used to define the law of the system dynamics.

Besides integral principles, differential (local by time) ones are used, for example, Gauss's principle of least constraint, the principle of least dissipation of energy, etc. According to M. Planck [33], local principles have some advantage over integral ones because the current state and motion of the system do not depend on its later states and motions there. Using [10], [14], let us formulate one more local variational principle based on the speed-gradient method proposed earlier to synthesize the laws of the nonlinear and adaptive control.

The speed-gradient principle: *of all the probable motions, the system implements the ones for which input variables vary directly as the speed-gradient of some "goal" functional Q_t .* If restrictions are imposed on a motion of the system then the direction is a speed-gradient vector projection on the acceptable directions (the ones that satisfy the restrictions) set.

The SG principle can be represented in the finite form

$$u = -\Gamma \nabla_u \dot{Q}_t. \quad (3)$$

where \dot{Q}_t is a rate of change of the goal functional along the trajectory of the system (2). Let us describe application of the SG principle in the simplest (but at the same time in the most important) case where a category of models of the dynamics (2) is specified as the relation:

$$\dot{x} = u. \quad (4)$$

The relation (4) just means that we are deriving the law of variation of variables speeds of the system state. In accordance with the SG principle, the goal functional (function) $Q(x)$ needs to be specified first. $Q(x)$ should be based on physics of a real system and reflect its tendency to decrease the current $Q(x(t))$ value. After that, the law of dynamics can be expressed as (3).

The SG principle is also applicable to develop models of the dynamics of distributed systems that are described on infinite-dimensional state spaces. There in particular can be a vector x in a Hilbert space \mathcal{X} and a nonlinear operator $f(x, u, t)$ defined on a dense set $D_F \subset \mathcal{X}$; in this case, the solutions of equation (2) are generalized.

III. JAYNES'S MAXIMUM ENTROPY PRINCIPLE

The approach proposed by Jaynes [2]–[4] became the foundation for statistical physics nowadays. Its main ideas are described below.

Let $p(x)$ be a probability density function (pdf) of a multidimensional random variable x . This is an unknown function that needs to be defined on the basis of certain system information. Let us suppose

that there is the information about some average values \overline{H}_m which are known a priori:

$$\overline{H}_m = \int H_m(x)p(x)dx, \quad m = 1, \dots, M. \quad (5)$$

The next equality is also true for the density function

$$\int p(x)dx = 1. \quad (6)$$

Conditions (5) and (6) in general can be insufficient to derive $p(x)$. In this case, according to Jaynes, applying maximization of information entropy S_I is the most objective method to define the density function.

$$S_I = - \int p(x) \log p(x)dx.$$

Maximum search with additional conditions (5) and (6) is performed by using Lagrange multipliers; it leads to

$$p(x) = \frac{1}{Z} \exp \left(- \sum_{m=1}^M \lambda_m H_m \right), \quad (7)$$

$$Z = \int \exp \left(- \sum_{m=1}^M \lambda_m H_m \right) dx, \quad (8)$$

where λ_m can be derived from conditions (5).

These formulas allow definition of the distribution function for microcanonical, canonical and other ensembles. We substitute the conditions characterizing each equilibrium ensemble for condition (5) [3]. In case of equilibrium (when appropriate random variables x are selected), the formulas also show that the maximum information entropy coincides with the Gibbs entropy and can be identified with the thermodynamic entropy.

Jaynes showed a close connection and succession between his approach and both the classical papers dealing with probability theory and statistics (Bernoulli, Laplace) and the papers doing with physics and information theory (in particular J. Gibbs and C. Shannon) [34].

Although information theory was initially based on some concepts of statistical physics, now, thanks to C. Shannon, the information approach can be taken as a basis to develop statistical physics. According to [1], formalism of statistical mechanics is some sequence of operations. Having limited information about microworld but following this sequence, we can get the most objective estimation (this is a method of error prevention).

IV. MAXIMIZATION OF ENTROPY WITH SPEED-GRADIENT METHOD

The maximum entropy principle of Gibbs and Jaynes [3], [4] defines the asymptotic behavior of the system, but does not say anything about how there is a movement to an asymptotic behavior. To answer this we use the SG principle.

A. System with a continuous distribution of states

Consider a system with a continuous distribution of the set of possible states. Probability distribution over states is characterized by pdf $p(t, r)$ which is continuous everywhere except for a set with zero measure:

$$\int_{\Omega} p(t, r) dr = 1, \quad (9)$$

where Ω is a compact carrier.

State of the system evolves over time. We are interested in the behavior of both steady and transient mode. Steady state is determined by the MaxEnt principle. If nothing else about the system is known then its supreme behavior will maximize its measure of uncertainty (entropy). As a measure of uncertainty we choose differential entropy which is defined as

$$S = - \int_{\Omega} p(t, r) \log p(t, r) dr. \quad (10)$$

Let's define a law of system dynamics as

$$\dot{x} = u(t, r), x = p(t, r). \quad (11)$$

We have to define a function $u(t, r)$.

According to speed-gradient principle a rate of entropy change (10) which is based on a system (11) has to be calculated first. Then a gradient of rate for function u has to be found. And finally, control parameters has to be defined. These parameters are proportional to projection of gradient on a surface of bounds (9).

Let's calculate \dot{S} :

$$\dot{S} = - \int_{\Omega} (u \log p + u \frac{p}{p}) dr = - \int_{\Omega} u \log p dr - \int_{\Omega} u dr. \quad (12)$$

From (9) it follows that

$$\int_{\Omega} u(t, r) dr = 0. \quad (13)$$

So, $\dot{S} = - \int_{\Omega} u \log p dr$. Gradient of \dot{S} by u is equal to $\nabla_u \dot{S} = - \nabla_u \langle \log p, u \rangle$. According to (50) we have that $\nabla_u \dot{S} = - \log p$.

Speed-gradient principle of motion forms $u = -\Gamma \log p(t, r) + \lambda$, where Γ can be taken as a scalar value and LaGrange multiplier λ is selected to satisfy a restriction (13).

$$\int_{\Omega} (-\Gamma \log p(t, r) + \lambda) dr = 0 \Rightarrow \lambda = \frac{\Gamma \int_{\Omega} \log p(t, r) dr}{\text{mes}(\Omega)}, \quad (14)$$

where

$$\text{mes}(\Omega) = \int_{\Omega} 1 d\Omega. \quad (15)$$

Final system dynamics equation has the following form

$$\dot{p} = -\Gamma \log p(t, r) + \frac{\Gamma \int_{\Omega} \log p(t, r) dr}{\text{mes}(\Omega)} = -\Gamma \left(\log p(t, r) - \frac{\int_{\Omega} \log p(t, r) dr}{\text{mes}(\Omega)} \right). \quad (16)$$

Eq. (16) can be represented in more general form

$$\dot{p} = -\Gamma(I - \Psi) \log p(t, r), \quad (17)$$

where $\Psi = \frac{\int_{\Omega} (\cdot) dr}{\text{mes}(\Omega)}$ is a linear operator which is invariant for p and I is an identity operator.

Physical meaning of equation (16) can be described as a movement to the direction of maximum rate of entropy production. It corresponds to MEPP.

1) *Equilibrium stability:* Let's investigate a stability of obtained equilibrium equation (16).

Theorem 1 *There exists a unique limit pdf $p^*(t, r)$ defined by equation (16):*

$S(p^(t, r)) = S_{max}$. And $p^*(t, r) = \text{mes}^{-1}(\Omega)$.*

Proof

Consider an entropy function of Lyapunov.

$$V(p) = S_{max} - S(p) \geq 0. \quad (18)$$

Let's find a derivative of function (19)

$$\dot{V}(p) = -\dot{S}(p) = \int_{\Omega} u \log p dr. \quad (19)$$

After substitution of expression for u from (16) to (19) we obtain

$$\dot{V}(p) = \int_{\Omega} -\Gamma \left(\log^2 p(t, r) - \frac{\int_{\Omega} \log p(t, r) dr}{\text{mes}(\Omega)} \log p(t, r) \right) dr.$$

After several conversions it is true that

$$\dot{V}(p) = -\frac{\Gamma}{\text{mes}(\Omega)} \left(\int_{\Omega} \text{mes}(\Omega) \log^2 p(t, r) dr - \left(\int_{\Omega} \log p(t, r) dr \right)^2 \right).$$

We will use the Cauchy-Bunyakovsky inequality

$$\left| \int_{\Omega} f(x) g(x) dx \right|^2 \leq \left(\int_{\Omega} |f(x)|^2 \right) \left(\int_{\Omega} |g(x)|^2 \right) \quad (20)$$

for functions $f = \log p$ and $g = 1$. Taking into account that a scalar value Γ is positive we get that $\dot{V}(p) \leq 0$. Let's denote a set of functions where \dot{V} has null values as $D = \{p(r, t) : \dot{V} = 0\}$. It is known that equality in the Cauchy-Bunyakovsky inequality is achieved when multiplicity occurs, i.e. $f(x) = \alpha g(x)$. In our case $\dot{V}(p) = 0$ is true when $\log p(t, r) = \alpha$. It is possible only when $p(t, r) = \text{const}$. Using restriction (9) we get that $\text{const} = \text{mes}^{-1}(\Omega)$. And set D consists of only one solution $D = \{p^*\}$.

□

2) *Asymptotic convergence:* We will show an asymptotic convergence of all solutions to p^* . To do it we use Barbalat's lemma.

Barbalat's lemma *If differentiable function $f(t)$ has a finite limit for $t \rightarrow \infty$ and its derivative $\dot{f}(t)$ is uniformly continuous then $\dot{f}(t) \rightarrow 0$ for $t \rightarrow \infty$*

Theorem 2 *For all pdfs defined by equation (16) it is true that $p(t, r) \rightarrow p^*$ for $t \rightarrow \infty$.*

Proof

For the sake of simplicity we define a designation for V in (18) as $v(t) = V(p(t))$. We will use a Barbalat's lemma to show that $\dot{v}(t) \rightarrow 0$. We use $v(t)$ as a function $f(t)$. Because of $v(t) \geq 0$ and $\dot{v} \leq 0$ the function $v(t)$ has a finite limit for $t \rightarrow \infty$.

Let's show conditions when function $|\dot{v}(t)|$ is bounded. That will lead us to the fact that \dot{v} is uniformly continuous.

$$\begin{aligned} \ddot{v}(t) &= -\frac{2\Gamma}{\text{mes}(\Omega)} \left(\text{mes}(\Omega) - 2 \int_{\Omega} \log^2 p dr \right) \int_{\Omega} \log p \frac{\dot{p}}{p} dr = \\ &= \frac{2\Gamma^2}{\text{mes}(\Omega)} \left(\text{mes}(\Omega) - 2 \int_{\Omega} \log^2 p dr \right) \left[\int_{\Omega} \frac{\log^2 p}{p} dr - \int_{\Omega} \frac{\log p}{p} dr \frac{\int_{\Omega} \log p dr}{\text{mes}(\Omega)} \right]. \end{aligned} \quad (21)$$

Expression $\int_{\Omega} \log^2 p dr < \infty$ comes true when

$$\text{mes}\{r : p(r) = 0 \cup p(r) = \infty\} = 0, \quad (22)$$

where mes is defined as (15).

A similar conclusion may be performed for function $\dot{v}(t)$. In this way a function \dot{v} is uniformly continuous if condition (22) is satisfied. Then according to Barbalat's lemma $\dot{v}(t) \rightarrow 0$ for $t \rightarrow \infty$.

Let $f_t = \log p_t$. Taking into account that $\int_{\Omega} f_t^2 dr = \|f\|^2$ the expression for \dot{v} may be written as

$$\begin{aligned} \dot{v} &= -\frac{\Gamma}{\text{mes}(\Omega)} [\text{mes}(\Omega) \|f_t\|^2 - (1, f_t)^2] = -\Gamma \|f_t\|^2 \left[1 - \frac{(1, f_t)^2}{\text{mes}(\Omega) \|f_t\|^2} \right] = \\ &= -\Gamma \|f_t\|^2 (1 - \cos^2(\alpha(t))), \end{aligned} \quad (23)$$

where $\alpha(t) = \cos(\langle 1, f_t \rangle)$.

If $\|f_t\|^2 \rightarrow 0$ then $\int_{\Omega} \log^2 p_t dr \rightarrow 0$. It means that $\text{mes}_t\{r : p_t(r) \neq 1\} \rightarrow 0$. It is equivalent to $\int_{\Omega} p_t dr \rightarrow \text{mes}(\Omega)$. Due to the bound for the density for all $\forall t$ it is true that $\int_{\Omega} p_t dr = 1$. Thus $\|f_t\|^2 \rightarrow 0$ is wrong if only $\text{mes}(\Omega) \neq 1$. The case when $\text{mes}(\Omega) = 1$ we consider as a degenerate. Given (23) we obtain $\alpha(t) \rightarrow 0$. It means that $\hat{f}_t \rightarrow \hat{1}$, where \hat{f}_t and $\hat{1}$ are normalized values for f_t and 1 respectively.

It follows that p_t tends to the stationary distribution. As explained earlier this distribution is unique.

Thus, $p_{t \rightarrow \infty} \rightarrow p^*$. \square

3) *Convergence of differential entropies:* Let's prove the convergence of differential entropies

Theorem 3 *For differential entropy defined as*

$$H(t) = H(p(t, r)) = - \int_{\Omega} p(t, r) \log p(t, r) dr$$

it is true that

$$H(t)_{t \rightarrow \infty} \rightarrow H(p^*), \quad (24)$$

where pdf $p(t, r)$ is defined by equation (16) and p^ is limiting pdf: $p(t, r)_{t \rightarrow \infty} \rightarrow p^*(r)$.*

Proof

In [17] Godavarti and Hero provide sufficient conditions for convergence of differential entropies $H(p_n) \rightarrow H(p_*)$. The only assumptions are that $p_n \rightarrow p_*$ and that exists a constant L and some $k > 1$ such that for all N

$$\int p_N |\log p_N|^k \leq L \quad (25)$$

and

$$\int p |\log p|^k \leq L \quad (26)$$

For all t our pdf $p(t, r)$ from (16) the first assumption $p_n \rightarrow p_*$ is satisfied due to Theorem 2. Conditions (25) and (26) are also satisfied on the compact carrier Ω . According to [17] it means that (24) is true. \square

So with eq. (37) we are able to predict the evolution of system's entropy.

B. Total energy constraint

Restriction (9) can be interpreted as the law of conservation of mass on a space Ω . Let's consider a system with additional constraint for the law of total energy conservation. We will consider a conservative case when energy does not depend on a time. The new constraint may be described as

$$\int_{\Omega} p(t, r) h(r) dr = E, \quad (27)$$

where E is a common energy of a system and $h(r)$ is a density of energy.

Let's consider a system

$$\dot{p} = u. \quad (28)$$

The problem is to find an operator u which satisfies to both constraints (9) and (27) at any time t . And the following goal condition is true.

$$S(p(t, r))_{t \rightarrow \infty} \rightarrow S_{max}.$$

To solve this problem we use the speed-gradient method. As a goal function we take

$$Q(p) = S_{max} - S + \lambda'_1 \left(\int_{\Omega} p(t, r) h(r) dr - E \right) + \lambda'_2 \left(\int_{\Omega} p(t, r) dr - 1 \right),$$

where S is a differential entropy, λ'_1 and λ'_2 are LaGrange multipliers. Due to the speed-gradient principle an operator u has to be described as

$$u = -\Gamma \nabla_u \dot{Q}(p, u, t).$$

Let's calculate the derivative of a goal functional

$$\begin{aligned} \dot{Q} = \int_{\Omega} \dot{p}(t, r) \log p(t, r) dr + \int_{\Omega} \dot{p}(t, r) dr + \lambda'_1 \left(\int_{\Omega} \dot{p}(t, r) h(r) dr \right) + \lambda'_2 \int_{\Omega} \dot{p}(t, r) dr = \\ \int_{\Omega} u(t, r) \log p(t, r) dr + \lambda'_1 \int_{\Omega} u(t, r) h(r) dr + (\lambda'_2 + 1) \int_{\Omega} u(t, r) dr. \end{aligned}$$

Taking the gradient by u we get

$$\nabla_u \dot{Q} = \log p(t, r) + \lambda'_1 h(r) + (\lambda'_1 + 1) u(t, r).$$

Now then

$$u = -\Gamma \log p(t, r) + \lambda_1 h(r) + \lambda_2, \quad (29)$$

where $\lambda_1 = -\Gamma \lambda'_1$, $\lambda_2 = -\Gamma(\lambda'_2 + 1)$.

Now we find LaGrange multipliers λ_1 and λ_2 based on boundary conditions. For the sake of simplicity we omit the arguments of functions. So, for example we use $\log p$ instead of $\log p(t, r)$. It follows from the condition (13) that

$$-\Gamma \int_{\Omega} \log p dr + \lambda_1 \int_{\Omega} h dr + \lambda_2 \text{mes}(\Omega) = 0. \quad (30)$$

Condition (27) is equivalent to $\int_{\Omega} u h dr$. From this condition it should be

$$-\Gamma \int_{\Omega} \log(p) h dr + \lambda_1 \int_{\Omega} h^2 dr + \lambda_2 \int_{\Omega} h dr. \quad (31)$$

Solving the system of equations (30) and (31) we obtain

$$\lambda_1 = \Gamma \frac{\text{mes}(\Omega) \int_{\Omega} \log(p) h dr - \int_{\Omega} \log(p) dr \int_{\Omega} h dr}{\text{mes}(\Omega) \int_{\Omega} h^2 dr - \left(\int_{\Omega} h dr \right)^2}, \quad (32)$$

$$\lambda_2 = \Gamma \frac{\int_{\Omega} \log(p) dr \int_{\Omega} h^2 dr - \int_{\Omega} h dr \int_{\Omega} \log(p) h dr}{\text{mes}(\Omega) \int_{\Omega} h^2 dr - \left(\int_{\Omega} h dr \right)^2}. \quad (33)$$

Equations (32) and (33) are defined when denominator in both fractions is not equal to zero. If we take in the Cauchy-Bunyakovsky inequality $f = h$ and $g = 1$ then the following inequality comes true

$$\left| \int_{\Omega} h dr \right|^2 \leq \text{mes}(\Omega) \int_{\Omega} h^2 dr, \quad (34)$$

This inequality becomes equality when $h = \text{const}$. It means that all energy levels coincide. This case is supposed to be degenerate and not considered here. Thus the expression

$$\text{mes}(\Omega) \int_{\Omega} h^2 dr \neq \left(\int_{\Omega} h dr \right)^2. \quad (35)$$

is always true.

Given (32), (33) and substituting (29) in equation (28) we obtain the common equation of system dynamics:

$$\dot{p}(t, r) = -\Gamma \log p(t, r) + \Gamma \frac{\text{mes}(\Omega) \int_{\Omega} \log(p) h dr - \int_{\Omega} \log(p) dr \int_{\Omega} h dr}{\text{mes}(\Omega) \int_{\Omega} h^2 dr - \left(\int_{\Omega} h dr \right)^2} h(r) + \Gamma \frac{\int_{\Omega} \log(p) dr \int_{\Omega} h^2 dr - \int_{\Omega} h dr \int_{\Omega} \log(p) h dr}{\text{mes}(\Omega) \int_{\Omega} h^2 dr - \left(\int_{\Omega} h dr \right)^2}. \quad (36)$$

Law of evolution (36) can be represented in abbreviated form:

$$\dot{p} = -\Gamma(I - \Psi) \log p, \quad (37)$$

where I is an identity operator, Ψ is a linear integral operator that is independent of p

$$\Psi = \frac{(1, \cdot)}{\text{mes}(\Omega)} + \frac{\tilde{h}(\tilde{h}, \cdot)}{\|h\|^2 - \frac{1}{\text{mes}(\Omega)} (1, h)^2}, \quad (38)$$

$$\tilde{h} = h - \frac{1}{\text{mes}(\Omega)} \int_{\Omega} h dr.$$

1) *Stability of equilibrium:* Further we examine the equilibrium of obtained equation (36).

Theorem 4 *There exists a unique limit pdf p^* defined by equation (36): $S(p^*) = S_{max}$.*

$p^* = C e^{-\mu h}$, where $\mu = -\frac{\lambda_1}{\Gamma}$, $C = e^{\frac{\lambda_2}{\Gamma}}$ and λ_1, λ_2 are defined in (29).

Proof

Let's calculate the derivative of functional (18) for the system (36):

$$\begin{aligned} \dot{V} &= \int_{\Omega} \dot{p} \log p dr = \int_{\Omega} (-\Gamma \log p + \lambda_1 h + \lambda_2) \log p dr = \\ &= -\Gamma \int_{\Omega} \log^2 p dr + \lambda_1 \int_{\Omega} h \log p dr + \lambda_2 \int_{\Omega} \log p dr = \\ &= -\Gamma \int_{\Omega} \log^2 p dr + \lambda_1 \int_{\Omega} h \log p dr + \frac{\Gamma \int_{\Omega} \log p dr - \lambda_1 \int_{\Omega} h dr}{\text{mes}(\Omega)} \int_{\Omega} \log p dr = \\ &= \frac{\Gamma}{\text{mes}(\Omega)} \left[\left(\left(\int_{\Omega} \log p dr \right)^2 - \text{mes}(\Omega) \int_{\Omega} \log^2 p dr \right) + \frac{(\text{mes}(\Omega) \int_{\Omega} h \log p dr - \int_{\Omega} h dr \int_{\Omega} \log^2 p dr)^2}{\text{mes}(\Omega) \int_{\Omega} h^2 dr - \left(\int_{\Omega} h dr \right)^2} \right]. \end{aligned}$$

It can be proved that

$$\dot{V}(p) \leq 0. \quad (39)$$

For the proof of inequality (39) we define a functional:

$$\langle \cdot, \cdot \rangle : L_2(\Omega) \times L_2(\Omega) \rightarrow \mathbb{R}, \forall f, g \in L_2(\Omega)$$

$$\langle f, g \rangle = \text{mes}(\Omega) \int_{\Omega} f g dr - \int_{\Omega} f dr \int_{\Omega} g dr. \quad (40)$$

New functional has several useful properties

1. Linearity in the first argument

$$\forall f, g, h \in L_2(\Omega), \forall \lambda \in \mathbb{R} \quad \langle \lambda f + g, h \rangle = \langle \lambda f, h \rangle + \langle g, h \rangle.$$

2. Symmetry

$$\forall f, g \in L_2(\Omega) \quad \langle f, g \rangle = \langle g, f \rangle.$$

3. Positiveness and the condition of zero value

$\forall f \in L_2(\Omega) \quad \langle f, f \rangle \geq 0, \quad \langle f, f \rangle = 0 \Leftrightarrow f = \mu = \text{const.}$

Let's prove inequality (39) base on properties 1-3.

Obvious that for any $f, g \in L_2(\Omega)$ and $\lambda \in \mathbb{R}$ it is true that $f - \lambda g \in L_2(\Omega)$. This function has property

3: $\langle f - \lambda g, f - \lambda g \rangle \geq 0$. Using properties 1 and 2 we get the quadratic inequality with respect to λ :

$$0 \leq \langle f - \lambda g, f - \lambda g \rangle = \lambda^2 \langle g, g \rangle - 2\lambda \langle f, g \rangle + \langle f, f \rangle.$$

This inequality holds for any real λ . Hence the discriminant can not be positive.

$$D = 4\langle f, g \rangle^2 - 4\langle f, f \rangle \langle g, g \rangle.$$

Thus

$$\langle f, g \rangle^2 \leq \langle f, f \rangle \langle g, g \rangle. \quad (41)$$

If the equality takes place in (41) then there exists a unique solution $\lambda \in \mathbb{R}$ of an equation $\langle f - \lambda g, \langle f - \lambda g \rangle = 0$. But then by property 3 we have $\exists \mu \in \mathbb{R} : f - \lambda g = \mu \mathbf{1}$.

Substituting $f = \log p, g = h$ to the inequality (41) we get

$$\left(\text{mes}(\Omega) \int_{\Omega} h \log p dr - \int_{\Omega} h dr \int_{\Omega} \log p dr \right)^2 \leq \left(\text{mes}(\Omega) \int_{\Omega} \log^2 p dr - \left(\int_{\Omega} \log p dr \right)^2 \right) \left(\text{mes}(\Omega) \int_{\Omega} h^2 p dr - \left(\int_{\Omega} h dr \right)^2 \right).$$

Which implies the inequality (39).

Note that the equality (39) holds if and only if

$$\exists \lambda, \mu \in \mathbb{R} : \log p = \lambda h + \mu. \quad (42)$$

In the case of equilibrium the expression (29) can be written as

$$\log p(t, r) = \frac{\lambda_1 h(r)}{\Gamma} + \frac{\lambda_2}{\Gamma},$$

which coincides with (42). Thus there is a unique pdf corresponding to the equilibrium state:

$$p^* = C e^{-\mu h}, \quad (43)$$

where $\mu = -\frac{\lambda_1}{\Gamma}$, $C = e^{\frac{\lambda_2}{\Gamma}}$. This state corresponds to the Gibbs distribution, which is consistent with the results of classical thermodynamics. \square

2) *Asymptotic convergence:* We prove asymptotic convergence with Theorem 5.

Theorem 5 For all pdfs defined by equation (36) it is true that $p(t, r) \rightarrow p^*$ for $t \rightarrow \infty$.

Proof

The proof of asymptotic convergence for eq. (36) will be performed similar to the case with one constraint (Theorem 2). To use Barbalat's lemma we have to check conditions under which the function \ddot{v} is limited.

$$\dot{v} = \frac{\Gamma}{\text{mes}(\Omega)} \left(\left(\int_{\Omega} \log p dr \right)^2 - \text{mes}(\Omega) \int_{\Omega} \log^2 p dr \right) + \frac{\Gamma}{\text{mes}(\Omega)} \frac{(\text{mes}(\Omega) \int_{\Omega} h \log p dr - \int_{\Omega} h dr \int_{\Omega} \log p dr)^2}{\text{mes}(\Omega) \int_{\Omega} h^2 dr - (\int_{\Omega} h dr)^2}. \quad (44)$$

We obtain

$$\ddot{v} = \frac{2\Gamma}{V} \left(\left(\int_{\Omega} \log p dr \int_{\Omega} \frac{\dot{p}}{p} dr - \text{mes}(\Omega) \int_{\Omega} \log p \frac{\dot{p}}{p} dr \right) + \frac{2\Gamma}{V} \left(\frac{(\text{mes}(\Omega) \int_{\Omega} h \frac{\dot{p}}{p} dr - \int_{\Omega} h dr \int_{\Omega} \frac{\dot{p}}{p} dr) (\text{mes}(\Omega) \int_{\Omega} h \log p dr - \int_{\Omega} h dr \int_{\Omega} \log p dr)}{V \int_{\Omega} h^2 dr - (\int_{\Omega} h dr)^2} \right) \right). \quad (45)$$

Considering (36) we get that \ddot{v} is limited under the same conditions which were used in the case of one restriction.

Namely, $\text{mes}\{r : p(r) = 0 \cup p(r) = \infty\} = 0$.

According to Barbalat's lemma it is true that

$$\dot{v} \rightarrow 0. \quad (46)$$

We introduce a scalar product as $\langle f, f \rangle = \text{mes}(\Omega) \int_{\Omega} f^2 dr - (\int_{\Omega} f dr)^2$. Having $\|f\|^2 = \langle f, f \rangle$ the expression (44) can be written as

$$\dot{v} = \frac{\Gamma}{\text{mes}(\Omega)} \left[-\|\log p\|^2 + \frac{(\log p, h)^2}{\|h\|^2} \right] = -\frac{\Gamma}{\text{mes}(\Omega)} \|\log p\|^2 \left[1 - \frac{(\log p, h)^2}{\|\log p\|^2 \|h\|^2} \right]. \quad (47)$$

Consider the case when $\|\log p\|^2 \rightarrow 0$. By Cauchy-Bunyakovsky inequality the equality in inequality

$$\left(\int_{\Omega} \log p dr \right)^2 \leq \text{mes}(\Omega) \int_{\Omega} \log^2 p dr$$

takes place when $\log p = \alpha$.

We have previously demonstrated that the equality $\dot{v} = 0$ holds in the only one case when $\log p = \lambda h + \mu$, where λ and μ are constants. Then h must be a constant since the equality $\lambda h + \mu = \alpha$. Such a case we assume degenerate because then the denominator in the expressions for the coefficients λ_1 and λ_2 ((32) (33)) is equal to zero.

Given (46), (47) and $\|\log p\|^2 \not\rightarrow 0$ we obtain $\frac{(\log p, h)^2}{\|\log p\|^2 \|h\|^2} \rightarrow 1$. Which implies that $\widehat{\log p} \rightarrow \widehat{h}$, where $\widehat{\log p}$ and \widehat{h} are normalized values for $\log p$ and h respectively. Thus p tends to the only one stationary distribution p^* since h does not depend on time. \square

3) *Convergence of differential entropies*: Similar to differential entropies convergence described in paragraph 4.1.3 (see Theorem 3) the convergence of differential entropies for the case of two constraints can be proved.

Theorem 6 *For pdf which is defined by equation (36) it is true that*

$$H(t)_{t \rightarrow \infty} \rightarrow H(p^*),$$

where $H(t) = H(p(t, r))$ is differential entropy and p^* is limiting pdf: $p(t, r)_{t \rightarrow \infty} \rightarrow p^*(r)$.

V. CONCLUSION

The MEPP is widely used in different studies of complex systems of physical, chemical or biological origin [1]. Using MEPP, P. Zupanovic [35] derived Kirchhoff law of electrical circuits.

The task to justify MEPP was unresolved for a long time. We propose such a justification based on results originated in the control theory.

It is an important problem to derive the principle that determines the dynamics of non-stationary (transient) states, and describes a way and trajectory of the system that tends to the state with maximum entropy. If a goal function is the entropy of a system then the extreme SG principle supplements Gibbs and Jaynes maximum entropy principle to determine the direction of evolution of the system when it tends to the state of maximum entropy. This direction corresponds to maximum entropy production rate (growth rate).

SG principle generates equations for the transient (non-stationary) states of the system operation, i.e. it gives an answer to the question of **How the system will evolve?** This fact distinguishes the SG principle from the principle of maximum entropy, the principle of maximum Fisher information and others characterizing the steady-state processes and providing an answer to the questions of **To where?** and **How far?**

The equations derived on the basis of the SG principle allow forecasting the dynamics of non-equilibrium systems with continuous distribution of parameters. These equations can prove to be useful to study both evolutions and relaxations of non-equilibrium systems of macroscopic and microscopic world.

APPENDIX

Preliminary Information

Scalar Product. In space of real functions $L_2(\Omega)$ a scalar product can be specified as

$$\langle f, g \rangle = \int_{\Omega} f(x)g(x)d\Omega, \quad (48)$$

where Ω is a compact carrier. This scalar product generates a norm:

$$\|f\|_2 = \sqrt{\langle f, f \rangle}. \quad (49)$$

Allowing completeness of space $L_p(\Omega)$ we can consider space $L_2(\Omega)$ as a Hilbert space.

Variational Derivative and Gradient of Scalar Product. Let us assume that functional Φ is specified as $\Phi : L_2(\Omega) \rightarrow \mathbb{R}$. A variational derivative of Φ can be specified as the analog of a finite-dimensional gradient.

$$d\Phi = \left(\frac{d\Phi}{df}, df \right) = \int_{\Omega} \frac{d\Phi}{df}(x) df(x) dx,$$

where $\frac{d\Phi}{df}$ is a variational derivative of Φ with respect to f , and df is a variation. If variation $d\Phi = \Phi(f + df) - \Phi(f)$ can be expressed as $d\Phi = (A, df) = \int_{\Omega} A(x) df(x) dx$ then A is a variational derivative of Φ with respect to f .

Thus for each scalar product $(a, g) = \int_{\Omega} a(x)g(x)dx$ in space of real functions $L_2(\Omega)$ a function a can be regarded as a variational derivative with respect to g , i.e.

$$\nabla_g(a, g) = a. \quad (50)$$

Differential and information entropy. If discrete random variable X is x_1, x_2, \dots, x_n with probabilities p_1, p_2, \dots, p_n , respectively then its measure of uncertainty is

$$H(X) = - \sum_{i=1}^n p_i \log(p_i). \quad (51)$$

Let us consider $p(x)dx$ as a probability of random variable hitting into indefinitely small interval $(x, x + dx)$ then the entropy in (51) can be formalized for a continuous case as:

$$H(X) = - \int_{-\infty}^{\infty} (p(x)dx) \log(p(x)dx) = - \int_{-\infty}^{\infty} p(x) \log(p(x)) dx - \int_{-\infty}^{\infty} p(x) \log(dx) dx, \quad (52)$$

where $p(x)$ is a pdf.

In expression (52) only the first term depends on the probability density of a random variable but the second one is always ∞ . The measures of uncertainty for different distributions can be compared by comparing only the first terms of expression (52). To do it, a concept of differential measure of uncertainty (differential entropy) is introduced.

$$h(X) = - \int_{-\infty}^{\infty} p(x) \log(p(x)) dx.$$

Differential entropy as an entropy of continuous distribution can be interpreted as the average information of continuous source, i.e. information entropy.

Properties of functional $\langle f, g \rangle$ from the proof of Theorem 2 (sect. 4.2.2)

1. Linearity in the first argument

$$\forall f, g, h \in L_2(\Omega), \forall \lambda \in \mathbb{R} \quad \langle \lambda f + g, h \rangle = \langle \lambda f, h \rangle + \langle g, h \rangle.$$

Proof Using the linearity of the integral we obtain

$$\begin{aligned} \langle \lambda f + g, h \rangle &= \text{mes}(\Omega) \int_{\Omega} (\lambda f + g) h dr - \int_{\Omega} (\lambda f + g) dr \int_{\Omega} h dr = \\ &= \left(\text{mes}(\Omega) \int_{\Omega} \lambda f h dr - \int_{\Omega} \lambda f dr \int_{\Omega} h dr \right) + \left(\text{mes}(\Omega) \int_{\Omega} g h dr - \int_{\Omega} g dr \int_{\Omega} h dr \right) = \\ &= \langle \lambda f, h \rangle + \langle g, h \rangle. \quad \square \end{aligned}$$

2. Symmetry

$$\forall f, g \in L_2(\Omega) \quad \langle f, g \rangle = \langle g, f \rangle.$$

Proof

$$\langle f, g \rangle = \text{mes}(\Omega) \int_{\Omega} f g dr - \int_{\Omega} f dr \int_{\Omega} g dr = \text{mes}(\Omega) \int_{\Omega} g f dr - \int_{\Omega} g dr \int_{\Omega} f dr = \langle g, f \rangle. \quad \square$$

3. Positiveness and the condition of zero value

$$\forall f \in L_2(\Omega) \quad \langle f, f \rangle \geq 0, \quad \langle f, f \rangle = 0 \Leftrightarrow f = \mu = \text{const.}$$

Proof Let's consider a scalar multiplication (48). Cauchy-Bunyakovsky comes true for it: $|(f, g)|^2 \leq (f, f)(g, g)$. $|(f, g)|^2 = (f, f)(g, g) \Leftrightarrow \exists \mu \in \mathbb{R} : f = \mu g$. Substituting $g \equiv 1$ we get

$$\left| \int_{\Omega} f dr \right|^2 \leq (f, f) \int_{\Omega} 1 dr = \text{mes}(\Omega) \int_{\Omega} f^2 dr.$$

Thereby it is true that

$$\langle f, f \rangle = \text{mes}(\Omega) \int_{\Omega} f^2 dr - \left(\int_{\Omega} f dr \right)^2 \geq 0.$$

Moreover $\langle f, f \rangle = 0 \Leftrightarrow \left| \int_{\Omega} f dr \right|^2 = (f, f)(1, 1) \Leftrightarrow \exists \mu \in \mathbb{R} : f = \mu 1, \dots f = \mu \quad \square$

REFERENCES

- [1] L.M.Martyushev, V.D. Seleznev, "Maximum entropy production principle in physics, chemistry and biology" Phys. Report, V.426 (1), pp. 1–45, 2006.
- [2] Jaynes E.T., "The minimum entropy production principle" Ann. Rev. Phys. Chem, V. 31, pp. 579–601, 1980.
- [3] Jaynes E.T., "Information theory and statistical mechanics" Phys. Rev, V. 106. pp. 620–630, 1957;
- [4] Jaynes E.T., "Information theory and statistical mechanics.2" Phys. Rev, V. 108. pp. 171–190, 1957.
- [5] Dewar R., "Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states" J. Phys. A: Math. Gen, V. 36. pp. 631–641, 2003.
- [6] Dewar R., "Maximum entropy production and the fluctuation theorem" J. Phys. A: Math. Gen, V. 38, pp. 371–381, 2005.
- [7] Bellman R., Dreyfus S., "Applied Dynamic Programming Princeton" New Jersey, 1962.
- [8] Tsirlin A.M., Mironova V.A., Amelkin S.A., Kazakov V., "Finite-time thermodynamics: Conditions of minimal dissipation for thermodynamic process with given rate" Phys. Rev. E, V.58, pp. 215–223, 1998.
- [9] Fradkov A L., "Speed-gradient entropy principle for nonstationary processes" Entropy, Vol. 10(4), pp. 757-764, 2008.
- [10] Fradkov A.L., Miroshnik I.V., Nikiforov, V.O., "Nonlinear and Adaptive Control of Complex Systems" Kluwer Acad. Publ., 1999.
- [11] Fradkov A. L., "Application of cybernetic methods in physics" Physics-Uspekhi, V.48, pp. 103–127, 2005.

- [12] Fradkov A., Krivtsov A., "Speed-gradient principle for description of transient dynamics in systems obeying maximum entropy principle" Bayesian inference and maximum entropy methods in science and engineering, V. 1305, pp. 399–406, 2010.
- [13] Fradkov A. L., "Speed-Gradient scheme and its application in adaptive-control problems" Automation and Remote Control, Vol. 40(9), pp. 1333–1342, 1979.
- [14] Fradkov A. L., "Cybernetical physics: from control of chaos to quantum control" Springer-Verlag, 2007.
- [15] S.-W. Ho and R. W. Yeung, "The Interplay Between Entropy and Variational Distance" IEEE Trans. Inform. Theory, vol. 56, no. 12, pp. 5906–5929, Dec. 2010.
- [16] S.-W. Ho and R. W. Yeung, "On information divergence measures and a unified typicality" IEEE Trans. Inform. Theory, vol. 56, no. 12, pp. 5893–5905, Dec. 2010.
- [17] M. Godavarti, A. Hero, "Convergence of differential entropies" IEEE Trans. Inform. Theory, vol. 50, no. 1, pp. 171–176, Jan. 2004
- [18] C.-N. Chuah, D.N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in mimo wireless systems under correlated fading" IEEE Trans. Inform. Theory, vol. 48, pp. 637–650, Mar. 2002
- [19] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multipiereceive antenna architectures," IEEE Trans. Inform. Theory, vol. 48, pp. 3117–3128, Dec. 2002.
- [20] X. Mestre, J. R. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO channels: Asymptotic evaluation under correlated fading" IEEE J. Select. Areas Commun., vol. 21, pp. 829–838, June 2003.
- [21] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals" IEEE Trans. Inform. Theory, vol. 39, pp. 1728–1735, Sept. 1993.
- [22] L. Zheng and D. N. C. Tse, "Packing spheres in the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel" IEEE Trans. Inform. Theory, vol. 48, pp. 359–383, Feb. 2002.
- [23] A. Gersho, "Asymptotically optimal block quantization" IEEE Trans. Inform. Theory, vol. IT-25, pp. 373–380, July 1979.
- [24] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing" IEEE Trans. Inform. Theory, vol. IT-14, pp. 676–683, Sept. 1968.
- [25] A. Kuh and B.W. Dickinson, "Information capacity of associative memories" IEEE Trans. Inform. Theory, vol. 35, pp. 59–68, Jan. 1989.
- [26] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. van der Meulen, "Nonparametric entropy estimation: An overview" Int. J. Math. Stat. Sci., vol. 6, no. 1, pp. 17–39, June 1997.
- [27] I. A. Ahmad and P.-E. Lin, "A nonparametric estimation of the entropy for absolutely continuous distributions," IEEE Trans. Inform. Theory, vol. IT-22, pp. 688–692, May 1976
- [28] A. Mookkadem, "Estimation of the entropy and information of absolutely continuous random variables," IEEE Trans. Inform. Theory, vol. 35, pp. 193–196, Jan. 1989.
- [29] J. E. Shore, R. W. Johnson "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy" IEEE Trans. Inform. Theory, vol. 26, no. 1, pp. 26–37, Jan. 1980.
- [30] S. Amari and H. Nagaoka, "Methods of Information Geometry" American Mathematical Society, Oxford University Press, 2000
- [31] S. A. Ali, C. Cafaro, A. Giffin, D. -H. Kim, "Complexity Characterization in a Probabilistic Approach to Dynamical Systems Through Information Geometry and Inductive Inference" Physica Scripta, 85(2), Febr. 2012.
- [32] Lanczos K., "The Variational Principles of Mechanics" Toronto, 1962.
- [33] Planck M., "The unity of physical world views" Physikalische Zeitschrift, Vol.10, pp. 62–75, 1909.
- [34] Jaynes E.T., "Delaware Seminar in the Foundations of Physics" ed. by M. Bun-ge. Berlin: Springer-Verlag, 1967.
- [35] Zupanovic P., Juretic D., Botric S., "Kirchhoff's loop law and the maximum entropy production principle" Phys. Rev. E, V.70, pp. 056108, 2004.